



Enhanced MSER Algorithm for Text Extraction

Jaswanth P

*Department of CSE&IT
Saveetha School of Engineering
Saveetha University, Chennai, INDIA*

Anusuya S

*Department of CSE&IT
Saveetha School of Engineering
Saveetha University, Chennai, INDIA*

Anil Kumar M

*Department of CSE&IT
Saveetha School of Engineering
Saveetha University, Chennai, INDIA*

Dhikhi.T

*Department of CSE&IT
Saveetha School of Engineering
Saveetha University, Chennai, INDIA*

Abstract- Text extraction from natural scene images is a challenging problem due to the variations in color, font size, text alignment, illumination etc. And it is a technique to identify and isolate the desired text from the images. In this work, we propose a novel method referred as Enhanced Maximally Stable Extremal Region (EMSER) to extract the text present in the images. The existing approaches deal with the same are lacking in accuracy. The proposed Enhanced Maximally Stable Extremal Region (EMSER) algorithm works with morphological operators such as dilation and reconstruction with different size of structuring elements to identify the shape of the text objects and to find the number of connected components accurately. The proposed method has been compared with Sobel and Canny edge detection methods and the superiority of proposed method has been shown in experimental results.

Keywords- MSER, Canny, Sobel, EMSER, Image and Text Extraction.

I. INTRODUCTION

The embedded text in images contains valuable information and is exploited in many content-based image and video applications, such as content-based web image search, video information retrieval, mobile based text analysis and recognition. Due to complex background, the variations of font, size, color and orientation, it is difficult to extract the text accurately. The natural images particularly consist of many objects of difference sizes and shapes. The text extraction techniques find the image regions concentrating on the text objects of regular or irregular shapes for the given images. In some cases, text detection becomes even meaningful by itself. For example, finding the appearance of a caption in news video can help to locate the beginning.

Text detection in natural scene images is an important prerequisite for many content-based image analysis tasks. In this paper, we propose an accurate and robust method for detecting texts in natural scene images. A fast and effective pruning algorithm is designed to extract Maximally Stable Extremal Regions (MSERs) as character candidates using the strategy of minimizing regularized variations. Character candidates are grouped into text candidates by the single-link clustering algorithm, where distance weights and clustering threshold are learned automatically by a novel self-training distance metric learning algorithm. The posterior probabilities of text candidates corresponding to non-text are estimated with a character classifier; text candidates with high non-text probabilities are eliminated and texts are identified with a text classifier. Experiments on multilingual, street view, multi-orientation and even born-digital databases also demonstrate the effectiveness of the proposed method.

II. LITERATURE SURVEY

The growth of multimedia documents are increasing every day, information identification, and requirement and indexing are of mandatory requirement for the emerging domains. The method called gamma correction method has been applied on the images and its strength has been checked [1]. Morphological based thresholding has been applied to detect the text from the images [2]. Segmentation of characters from the image help to find the characters and it is compared with the standard methods [3]. Human-machine interactive software application was proposed to work with Chinese characters [4]. The connected component based approach is proposed and it is compared with recent approaches [5]. A statistical model based scheme and conditional random field (CRF) proposed for automatic extraction in born digital images [6] [7]. Text extraction applied to biomedical images [8]. An algorithm comprising the steps such as detection and localization, text enhancement



and segmentation and optical character recognition (OCR) proposed in natural scene images [9]. A novel text extraction technique is demonstrated for handwritten English or Bengali text [10]. An unsupervised feature selection is introduced for text-graphics segmentation problem [11]. Adaptive color reduction (ACR) followed by page layout analysis (PLA) approach implemented to automatically detect and extract text in mixed-type color documents [12]. A metric-based clustering is used to extract textual information from real-world images [13]. In this work we propose EMSER algorithm to work with images of different type, size and shape.

III. PROPOSED EMSER ALGORITHM

In this work, we propose a robust and accurate EMSER method to extract the text from images. This method, Enhanced MSER uses morphological operations for preprocessing. The dilation and reconstruction are used to enhance the images. The various structuring elements (strel) such as line and disk with different sizes are used. The region filling is used to fill the text to be isolated from the background. The flow of the work is mentioned in Figure 1.

The mapping of image is given as $I: D \subset Z^2 \rightarrow S$, Extremal regions are well defined if S is totally ordered and n adjacency relation is to be defined as $A \subset D \times D$

The regions are defined exclusively by the intensity function, this leads to identification of many key characteristics of the regions.

The input image is loaded and converted into gray image. MSER regions are found for the input image. All pixels inside the MSER have higher or lower intensities than in the surrounding regions and regions are selected to be stable over intensity range. The morphological operation dilation and reconstruction are used to improve the clarity of text and to find the connected components. The edges of connected regions are isolated using edge detection technique. The canny and sobel edge detection algorithms were applied on the resultant images and found that the performance of sobel is better than canny.

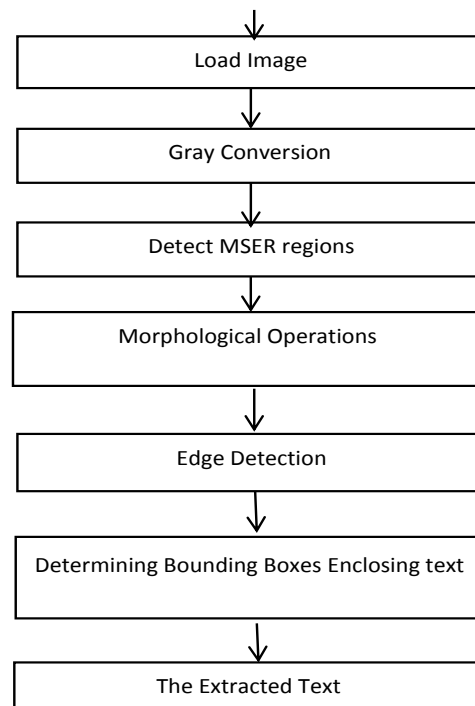


Figure 1. Architecture of proposed method.

The step by step procedure of the proposed method is given below:

i) Loading the input image

Load the image I with size $M \times N \times P$ where $x_{i,j,k}$ refers the color intensity $I = \sum \sum \sum x(i, j, k)$ where $i \rightarrow 1$ to $M, j \rightarrow 1$ to $N, k \rightarrow 1$ to P .



ii) *Converting to gray color Image*

Convert image I into gray scale image II where $x_{i,j}$ is the gray scale intensity $I_I = \sum \sum x(i, j)$ where $i \rightarrow 1$ to $M, j \rightarrow 1$ to N .

iii) *Image Enhancement*

Apply morphological operations image dilation and reconstruction with structuring element size 3×3 on the regions obtained in step 2.

iv) *Edge Detection*

Detect the edges present in the text region.

v) *Text Region Detection*

Detect regions R which is subset of $I_I, Q \subset I_I$.

vi) *MSER Detection*

Detect extremal regions $Q \subset R \subset I_I$ for all $p \in Q, q \in \partial Q$ and ∂Q is the outer region boundary.

Steps for Canny Edge detection


- a) Smooth the image with Gaussian filter
- b) Find intensity gradient of the image
- c) Remove low intensity pixels
- d) Apply threshold to find the edges
- e) Remove weak edges and highlight the strong edges

Sobel algorithm finds discrete differentiation by computing an approximation of the gradient of the image. The operation of Sobel operator is based on convolving the image with a small, separable, and integer valued filter in horizontal and vertical direction. The implementation results of Canny, Sobel and proposed algorithm are tabulated Table I.



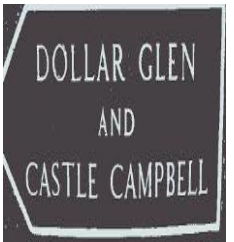


IV. EXPERIMENTAL RESULTS

In this section, the proposed method has been tested on different images and the accuracy was observed over other methods.



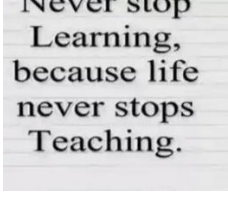

TABLE I. DIFFERENCE BETWEEN EXISTING AND EMSER METHODS FOR TEXT EXTRACTION

Image	Results using canny	Results using sobel	Results using EMSER
	's)'6H\\TJ*\Jl*_III\Jl _HHHHHHLJl'IH < @\>¢ A J; 5; cuv 'q/\ "w"4'è K > A 'R\$\$T£%:@:ED2_/WWW x L 'F o 'D 5"r& /r\.,gj; '\$/*"/	H'WH\H»'HH \HV HMH RE\$TAMLNT	RESTAURANT Connected components= 37



	<p>NC») LUJ[fD@ @jiEU1©\T‘_\$_@@ @m1fEwy _rm@ @</p> <p>Ems: %1_@@D f@@,Wmm_f: f:©</p> <p>W@{fPK_ \$y\$_@m mm ©@@r@_<@m3</p> <p>E@&@i’s@s M@ma@@Lm@Uuti \$@@G_@>m</p>	<p>;:/ E { N@ @@@@UU@€_\$_@@ ,@mfmj_UUfi©_hig W</p> <p>U @[?@\$fU)a X</p> <p>E°>@f_r_m_i: @@</p> <p>w©rrE@\ \$y\$!:@m Em @[@@Lr@_@_@m</p> <p>[E@Q@8©@ M@m1@mm[m8 \$@©R_@va</p>	<p>No unauthorised entry into this area!</p> <p>Permit to work system in operation</p> <p>Estates Management Section</p> <hr/> <p>Connected components= 320</p>
	<p>K ~: \ L9v@ @@@@ WEED? @ @ _n©cm°</p> <p>_x“ L%’J@_gh’& V</p> <p>wmm v@_\$_@©_</p>	<p>w_ith <:@LuJ~C:_©m \ lH]@_gjEmf:</p> <p>w[fn@tm [F@_\$_@ @1</p> <p>_5mm @@\$L”2”E</p> <p>E” W M</p>	<p>Proceed with caution. Height when raised</p> <p>115mm (4’2”) 1</p> <hr/> <p>Connected components= 188</p>
	<p>/ DOLLAR GLEN mm</p> <p>CASTLE CAMPEILL V’</p>	<p>// D©LLAIR GLEN AND</p> <p>“\C/ASTLE CAMP@ELJL “</p>	<p>‘DOLLAR GLEN AND CASTLE CAMPBELL V’</p> <hr/> <p>Connected components= 27</p>
		<p>AT % *1’ ; ESTATE AGENTS W’ XKQNSE</p> <p>Y\ \ 7 Na _ PWW» mm: .«H n</p>	<p>‘ ESTATE AGENTS MSAXONSI</p> <hr/> <p>Connected components= 88</p>
	<p>‘©© Ig©m3 U<m@m7 @mJy@m@ / \$i@vr&_im@ umW@v\$_uL _nk ytatm_ /E</p>	<p>/</p> <p>\©© :g@Iw Bf_w:m@w @my@m@</p> <p>&\$Kw_m@wm_%w\$_y # E_m\$g@Q_?</p>	<p>Do you Know anyone starting university \ his gear?</p> <hr/> <p>Connected components= 213</p>



	<p>©3@t_©m @ffU @m@@@§ ?% QL 7?d§>% \ \$0 9 % V ~% Mg \)1 %' s' R @@@m§ “ §©uth \nz cu \ ?!{»u\ \ V , » xmmx “NW »m\ W» 95\$ u A1_«\ \ J _F'X9M V\(\(\ M\ 33 n E xJ\ V fw (Q wm?_Nm>>,h<>wxmu\ mm\wm xxmw at</p>	<p>x \ “ n \ N n . \ UK' “ ”é©@ffri5_@@?@__©_ e -v a §©UiEh Ccarwgpug #</p>	<p>Accommodation Office . _____.) . South Cam o us Connected components= 260</p>
	<p>s.>,im.,:3ea <3 wmwvumm trannrmwam mmbamm:-HE I2::(:?,'21m- =m¢QL7;7</p>	<p>g,\u1mm«. E (rum ' mcsmw awn- rr 'k)uu',mm mm; @ u©</p>	<p>syllabus -80 GB_ We studv - 8o MB Retains in Mind » 8o KB write In Exams - 8o BYTES Result Comes In - BINARY DIG (. 3': Connected components= 45</p>
	<p>' g:«~ «:.,< if isaiése</p>	<p>P P, rL:~ „,—. K” V“ ’ ’* r_ I L' i kw</p>	<p>Never stop Learning, because life never stops Teaching. Connected components= 446</p>
	<p>VW_E9_@</p>	<p>N/My 3 §@:~L, , % r;J'€' LLW/K\M\vj\$V V@7U?R@</p>	<p>\VIPRO Connected components= 8</p>

The above table results show the differences between the canny, sobel and EMSER algorithms for text extraction. The connected components of EMSER represents the number of components with closed boundaries and have been mentioned in the Table I. Text extraction using EMSER is more effective, simple and user friendly.



TABLE II. THE PROCESS OF TEXT EXTRACTION USING THE PROPOSED EMSER

MSER regions filling	Edge detection using Sobel	Image enhancement using dilation	Text region extraction
			Proceed with caution. Height when raised 115mm (4'2")
Text region extraction from background	Connected components	Highlighting boundary of text	The output(extracted text)

Table II shows the results of our proposed method. The accuracy of the proposed and existing methods are checked using F-measure and is defined in (1).

TABLE III. COMPARISON OF PROPOSED METHOD WITH EXISTING METHODS

Method	Precision(P)	Recall(R)	F-measure
Sobel	0.45	0.32	0.37
Canny	0.33	0.37	0.34
Proposed	0.56	0.44	0.49

Precision P is defined as the ratio between areas of detected intersecting text regions, recall R is obtained from the ratio between area of intersection regions and that of ground truth regions.

$$f\text{- Measure} = (2 * \text{Precision} (P) * \text{Recall} (R)) / \text{Precision} (P) + \text{Recall} (R) \tag{1}$$

V. CONCLUSION

In this work, we have proposed a novel text extraction method referred as EMSER (Enhanced Maximally Stable Extremal Region) and it has been demonstrated on the images embedded with text and the accuracy obtained is 0.49 as shown in Table III. The proposed method works with the images of different sizes and different shapes with less computation cost. In future, this work can be extended to detect the text from video and can be automatically documented in Word Pad for further use.

REFERENCES

- [1] Devi. G.G. and Sumathi. C.P.,” Text extraction from images using gamma correction method and different text extraction methods — A comparative analysis” in Information Communication and Embedded Systems (ICICES), 2014 International Conference on 2014, DOI: 10.1109/ICICES.2014.7033973.
- [2] Hasan,Y.M.Y. and Karam,L.J., “Morphological text extraction from images”, Image Processing, IEEE Transactions on Volume:9, Issue:11, 2000, DOI: 10.1109/83.877220.
- [3] Basavaraj Amarapur and Nagaraj Patil, “Video Text Extraction from Images for Character Recognition”, Electrical and Computer Engineering, 2006. CCECE '06. Canadian Conference on 2006, DOI: 10.1109/CCECE.2006.277408.
- [4] Jian Yuan ; Yi Zhang ; Kok Kiong Tan ; Tong Heng Lee, “Text extraction from images captured via mobile and digital devices”, Advanced Intelligent Mechatronics, 2009. AIM 2009. IEEE/ASME International Conference on 2009, DOI: 10.1109/AIM.2009.5229951.
- [5] Talukder, K.H. and Mallick, T., “Connected component based approach for text extraction from color image” Computer and Information Technology (ICCIT), 2014 17th International Conference on 2014, DOI: 10.1109/ICCITechn.2014.7073114.



- [6] Roy, A., Parui, S.K. and Roy, U., “A Pair-Copula Based Scheme for Text Extraction from Digital Images” Document Analysis and Recognition (ICDAR), 2013 12th International Conference on 2013, DOI: 10.1109/ICDAR.2013.182.
- [7] Jian Zhang ; Renhong Cheng ; Kai Wang ; Hong Zhao ; Jiao Jiao, “Research on Born digital Image Text Extraction Based on Conditional Random Field”, P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), 2013 Eighth International Conference on 2013, DOI: 10.1109/3PGCIC.2013.62.
- [8] Songhua Xu and Krauthammer, M., “Boosting text extraction from biomedical images using text region detection”, Biomedical Sciences and Engineering Conference (BSEC), 2011, DOI: 10.1109/BSEC.2011.5872319.
- [9] Honggang Zhang; Kaili Zhao; Yi-Zhe Song; Jun Guo,” Text Extraction, Neuro computing, Volume 122, 25 December 2013.
- [10] S. Basu; C. Chaudhuri; M. Kundu ; M. Nasipuri ; D.K. Basu, ” text line extraction from multi-skewed handwritten documents”, Pattern Recognition, Volume 40, Issue 6, June 2007.
- [11] Pradipta Maji and Shaswati Roy, “Rough-fuzzy clustering and multiresolution image analysis for text graphics segmentation”, Applied Soft Computing, Volume 30, May 2015.
- [12] C. Strouthopoulos, N. Papamarkos and A.E. Atsalakis,” text extraction in complex color documents”, Pattern Recognition, Volume 35, Issue 8, August 2002.
- [13] Céline Mancas-Thillou and Bernard Gosselin, ” Color text extraction with selective metric-based clustering”, Computer Vision and Image Understanding, Volume 107, Issues 1–2, July–August 2007.